DOCUMENT RESUME

ED 261 077 TM 850 479

AUTHOR Frederick, Judith M.; Clauset, Karl H.
TITLE A Comparison of the Major Algorithms for Measuring School Effectiveness.
PUB DATE Mar 85
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Gains; *Educational Status Comparison; Elementary Education; Elementary Schools; *Evaluation Criteria; Evaluation Methods; Institutional Evaluation; Longitudinal Studies; Norms; Outcomes of Education; *Research Design; *Research Problems; *School Effectiveness; Time
IDENTIFIERS Aggregation (Data); Metropolitan Achievement Tests; Metropolitan Readiness Tests

ABSTRACT

Ten algorithms for measuring school effectiveness were identified through literature analysis and interviews of the directors of 39 school improvement projects. The algorithms are described thus: (1) a major proportion of students achieve at or above average national levels; (2) an equal percentage of highest and lowest social classes achieve minimum basic skills mastery; (3) percentage of average and high achievers is rising while percentage of low achievers is dropping; (4) proportion of low income students at minimum mastery is rising; (5) achievement gap of low achievers with respect to grade level is decreasing; (6) gaps between racial or socioeconomic status are stable or decreasing; (7) performance equals or exceeds city-wide norms; (8) two or more independent groups perform above the 75th percentile; (9) school mean gain equals or exceeds the city-wide gain; and (10) average achievement exceeds predicted mean achievement. Each algorithm is classified according to the time frame of analysis, the level of data aggregation, the continuity of the population, and the reference norms used. Six of these algorithms are applied to data collected from a suburban elementary school 30 miles from Boston on tests of reading comprehension, mathematics, and readiness. The differing results of the different algorithms are described. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED261077

A COMPARISON OF THE MAJOR ALGORITHMS

FOR MEASURING SCHOOL EFFECTIVENESS

BY

JUDITH M. FREDERICK

KARL H. CLAUSET

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. M. Frederick

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION
ANNUAL MEETING
MARCH 31 - APRIL 4, 1985
CHICAGO, ILLINOIS

TM 850 479

One of the greatest challenges facing educators today is to provide a quality education for all students. It is important for school practitioners and researchers to have the appropriate tools for measuring effectiveness to determine whether or not schools are meeting this challenge.

Literature reviews of the effective schools research reveal that there is no consensus on the definition of an effective school. The research is characterized by a variety of designs, methods, and measures of effectiveness making comparisons difficult. Educators involved in school improvement projects are faced with the dilemma of choosing from a variety of algorithms or "models" for measuring school effectiveness. Michael Kean (1982) warned, "unless the nature of effectiveness can be described and agreed upon, researchers face the possibility of identifying variables which may relate to the concept of an effective school not shared or accepted by those responsible for teaching children."

In developing a plan for measuring a school's effectiveness, the evaluator must consider such questions as:

*For whom should the school be effective?

*At what level should the data be aggregated?

*How will the variables be measured?

*Will the data be gathered at a single point in time or longitudinally?

It is our belief that there are appropriate answers to these questions and that educators who desire to implement school improvement projects and/or evaluate a school's performance can develop dependable, accurate tools with which to measure progress.

The State of the Art

The body of literature associated with the effective schools movement challenges the assumption that differences among schools have little or no impact on achievement. The last decade has witnessed a surge of research designed to demonstrate that schools do, in fact, have an impact on achievement.

Most researchers have attempted to define an effective school in terms of measurable student outcomes. The reasoning for this seems clear in that the quality of the "product" of the school, the student, is the most critical element of the effective school (Westbrook, 1982).

By and large, standardized achievement scores are utilized as the sole outcome measure in assessing a school's effectiveness in the existing literature. The use of standardized achievement tests as a measure of pupil performance and/or effectiveness has been the subject of considerable debate among educators for a number of years. It is generally agreed that the most important indicator of effectiveness is achievement and achievement gains. However, what level or gain denotes effectiveness and how it should be assessed remains unresolved. The research indicates that in practice the standards used have varied greatly.

Critics of the effective schools research have identified three major conceptual and methodological shortcomings in the research that has been conducted to date. First, there has been no systematic sampling of different types of schools. The existing body of research has concentrated on urban elementary schools. Miles (1983), in his review of thirty-nine school improvement programs, substantiates this criticism with data which indicates that while suburban areas comprise 30 percent of the communities of the United States, they are underrepresented in effective schools programs and studies.

Secondly, there is a paucity of longitudinal studies. For the most part, researchers have taken a "snapshot" of one year's achievement pattern. In their critical article, "Research on Effective Schools: A Cautionary Note", Rowan, Bossert, and Dwyer (1983) report that measures of effectiveness at the building level are not very consistent over time, although they are more consistent than would be expected by chance. Schools are often labeled as effective on the basis of instructional outcomes at only one or two grade levels. The use of algorithms or "models" which utilize data from a single point in time raises serious concerns about the stability of these measures. Few studies have required schools to be consistently effective in order to be described as effective.

Finally, the most typical means of gauging school effectiveness is through the use of aggregate scores of the pupils attending a particular school. The use of aggregate scores, generally in the form of a grade level or school mean, often masks critical differences among various cohort groups. Little or no attempt has been made to track the achievement patterns of individual students.

Method of Research

In response to the weaknesses identified in the literature, our study attempted to answer the following questions:

1. What are the major algorithms or models for measuring school effectiveness?
2. To what extent do the various algorithms ("models") lead to different conclusions about a school's effectiveness?
3. For each model, what is the stability of the effectiveness measure over time?
4. To what extent are the algorithms appropriate for use in suburban schools which are essentially homogenous in terms of race and socio-economic status?
5. What are the policy and program implications of using each algorithm in measuring school effectiveness?

Ten algorithms for measuring school effectiveness were identified through a process which involved a detailed analysis of the literature and phone interviews with the directors of thirty-nine school improvement projects. Each algorithm for measuring school effectiveness can be defined along four dimensions:

1. Time Frame of Analysis - Did the algorithm measure the effectiveness of the school at a single point in time or over a period of one or more years?
2. Level of Aggregation - Was the data collected and analyzed to produce school-wide averages or were specific cohort groups (based on SES, race, or achievement) identified and compared?
3. Continuity of Population - In using multi-year methods, were the scores for the same students compared over time or were scores for different groups of students compared?

4. Reference Norm - Is the determination of effectiveness based on internal standards (i.e. last year's performance) or an external standard (i.e. national norms)?

Figure 1 shows the results of this classification applied to the ten algorithms we identified.

The data examined in this study were collected from a suburban elementary school (k-6) located approximately thirty miles from Boston. The elementary school population consists largely of white middle class pupils. Approximately 15 percent of the students are identified as low income pupils on the basis of eligibility for free or reduced lunch.

Standardized achievement tests are administered annually in the subject school. In this study, two sub-tests scores of the Metropolitan Achievement Tests were used as the measure of effectiveness: Total Math and Reading Comprehension. Data from the Metropolitan Readiness Test were used for the algorithms which required cohort groups based on initial readiness.

Six of ten algorithms originally identified were investigated in this study. Three algorithms involving the comparison of school data to city-wide data were not applicable to data generated in a community which operates only one elementary school. A fourth algorithm involved a rank ordering of all

FIGURE 1

School Effectiveness Algorithms

ALGORITHM	TIME FRAME	LEVEL OF AGGREGATION	NORM	POPULATION
1. An effective school is one which provides quality education with a major proportion of the students achieving at or above the average national levels. (Brookover, 1981)	POINT IN TIME	SCHOOL	EXTERNAL (NATIONAL)	* *
2. An effective school is one in which an equal percentage of the highest and lowest social classes achieve minimum mastery in the basic skills. (Edmonds, 1982a;Connecticut, 1980)	POINT IN TIME	Cohorts based on Socio-economic Status	INTERNAL	* *
3. An effective school is one in which the percent of students scoring in the average and high achievement categories is increasing while the percent of students in low achievement category is dropping. (PROJECT R.I.S.E., 1979)	INTERVAL	Cohorts based on Achievement	INTERNAL	DIFFERENT
4. An effective school is one in which the proportion of low income students at minimum mastery is rising. (Edmonds, 1982b)	INTERVAL	Cohorts based on Socio-economic Status	INTERNAL	DIFFERENT
5. An effective school is one in which the achievement gap with respect to grade level standards for initially low achieving students is closed or is closing over time. (Clauset/Gaynor, 1982)	INTERVAL	Cohorts based on Initial Readiness	EXTERNAL (NATIONAL)	SAME
6. An effective school is one in which the initial gap between cohort groups based on race and socio-econmic status remains stable or is reduced. (MGAPb, 1981)	INTERVAL	Cohorts based on Race and Socio-economic Status	INTERNAL	SAME
7. An effective school is one which performs at or above the city-wide average. (Lezotte, 1974; MGAPa, 1981)	POINT IN TIME	SCHOOL	*EXTERNAL (CITY)	* *
8. An effective school is one in which 2 or more independent groups of pupils perform above the 75th percentile. (Frederickson, 1975)	POINT IN TIME	SCHOOL	EXTERNAL (NATIONAL)	* *
9. An effective school is one in which the school mean gain is at or above the city-wide mean gain. (N.Y.C.S.I.P., 1979)	INTERVAL	SCHOOL	*EXTERNAL (CITY)	DIFERENT
10. An effective school is one in which the observed average achievement exceeds the predicted mean achievement. (Salganik, 1980)	INTERVAL	SCHOOL	INTERNAL	SAME

* Quasi-external because the city mean is influenced by the school mean.

* * For Point-in-Time studies, "Same" or "Different" population has no meaning.

elementary schools within the district and thus was inappropriate for use in this study. Each algorithm was applied over three successive time intervals to estimate the stability of the effectiveness measures within each model.

RESULTS

The analysis of data revealed that the various algorithms for measuring school effectiveness do, in fact, lead to different conclusions depending upon the configuration of the four basic dimensions of the algorithm.

Two algorithms declared the school under study to be effective. The first algorithm (Brookover, 1981), which specifies that a major proportion of the pupils must achieve at or above the average national level, found that a major proportion of the students performed above the 50th percentile in both Reading and Mathematics for all three years. This algorithm uses a point-in-time measure with data aggregated at the school level.

The second algorithm to label the school as effectivewas proposed by Ronald Edmonds (1982b). In this algorithm, Edmonds defines an effective school as one in which the proportion of low income pupils at minimum mastery is rising. Analysis of the data for both Reading and Mathematics revealed that

the proportion of low income pupils above the 30th percentile (minimum mastery) increased over a four year period. For this analysis, the data was aggregated at the cohort level (based on SES) and gathered over an interval of time.

Only one algorithm (Clauset/Gaynor, 1982) demonstrated that the school was ineffective for the cohort of pupils identified as initially low achievers on the basis of their scores on the Metropolitan Readiness Test. This algorithm, which utilized longitudinal data gathered from Grade 1 through Grade 6, identifies an effective school as one in which the achievement gap with respect to grade level standards for initially low achieving students is closed or is closing over time. The data examined for this model revealed that the initial gap in achievement between the initially low achieving pupils and grade level standards actually widened over time for all three cohort groups. The gap in achievement widened from 4 months in the first grade to 2 years and 2 months in the sixth grade for the Class of 1987, from 4 months to 3 years for the Class of 1988, and from 4 months to 1 year and 6 months for the Class of 1989.

The three remaining algorithms examained in this study were inconsistent in their results. The first algorithm to produce inconsistencies defines an

effective school as one in which an equal percentage of the highest and lowest social classes achieve minimum mastery in the basic skills (Edmonds 1982a; Connecticut School Effectiveness Project). The data examined for this algorithm indicated that an equal percentage of the highest and lowest social classes achieved minimum mastery in Mathematics for all three years. However, in Reading the percentage was equal for only one of the three years. The data analyzed in this model were aggregated at the cohort level (based on SES) taken at a single point in time.

A second algorithm, utilized in Project R.I.S.E. (1979) also demonstrated inconsistent results. Project R.I.S.E. (1979) defines an effective school as one in which the percent of students scoring in the average and high achievement categories is increasing while the percent of students in the low achievement categories is dropping. The analysis of data revealed that in Grade 3 the overall percentage of pupils in the low achievement category (stanines 1, 2, 3) in Mathematics increased, while the percentage of pupils in the low achievement category (stanines 1, 2, 3) in both Mathematics and Reading experienced a decrease from 1980 - 1983. The data were analyzed according to

cohort groups (based on achievement) taken over a four year interval.

The final algorithm to produce inconsistent results states that an effective school is one in which the initial gap between cohorts based on race and socio-economic status remains stable or is reduced in Grades 1-6 (Middle Grades Assessment Program, 1981). The data analyzed for this algorithm revealed that the low income cohort scored above the average income cohort for the Class of 1987. In the area of Reading, the initial gap between the low income cohort and the middle income cohort increased for the Class of 1988 and decreased for the Class of 1989. The data revealed that the gap in Mathematics was completely closed for the initially low achieving cohort of the Class of 1988 and reduced by 8 percentage points for the Class of 1989. The data for this "model" included the achievement test data for Grades 1-6 aggregated at the cohort level (based on SES).

The results of this study and a similar study, conducted by Richard Silverman (1984) of Boston University, indicate that the algorithm developed by Clauset and Gaynor (1982) has proven to be more consistent than the other algorithms which analyze performance for different cohorts. This algorithm,

which is based on achievement cohorts whose performance is compared to grade level standards over a six year period, appears more promising than longitudinal algorithms based on SES or race. However, it has not been widely tested.

Methodological Issues

Several methodological issues have surfaced in the completion of this study, particularly in reference to the interpretation of data. It has become apparent that, in general, the algorithms lack sophistication in terms of the statistical procedures utilized in the models. The "models" ignore rudimentary statistical techniques which, if applied, would address the need for guidelines on how to interpret the data. For example, the results from the models which utilize interval or longitudinal data could be interpreted more effectively with the application of linear regression to determine the "line of best fit." The use of this technique would provide additional information to allow the educational decision-maker to determine whether the pattern of achievement is related solely to the previous achievement of the pupils or possibly to other factors.

A second area of concern surfaced in the interpretation of the algorithm which required that an equal percentage of low income anad middle income pupils achieve minimum mastery (Edmonds 1982a, Connecticut School Effectiveness Project, 1980). Since it was highly unlikely that the results for the two cohort groups would be identical, a range of plus or minus 5% was arbitrarily established as equal. The analysis of Reading data revealed a difference of 5% in 1981, 7% in 1982, and 6% in 1983. According to the criteria established initially, the school was classified as ineffective in 1982 and 1983 and the model was considered to be inconsistent in terms of stability, in spite of the fact that there was very little variance in the results. It would seem logical to further examine the data, in this instance, to determine the number of pupils in the low income cohort who are actually affected by the difference in the percentages. In 1982, 85 pupils achieved minimum mastery. If the results had equalled the criterion (5% difference), 87 pupils would have been included. The data for 1983 reveal a difference of only one pupil when the criterion of a five percent difference is used as the reference point. Thus, it appears that the results are actually relatively stable according to this additional

analysis.

A third issue which needs to be clearly addressed prior to the application of this algorithm (Edmonds, 1982a; Connecticut School Effectiveness Project, 1980) and one other (Edmonds, 1982b) is the criterion for minimum mastery. In this study, the thirtieth percentile was established as the level of minimum mastery in accordance with the guidelines established by Edmonds. This level was established in conjunction with Edmonds' work in urban schools. Educators who intend to use these algorithms as a measure of effectiveness must determine whether or not the thirtieth percentile is an appropriate criterion. Appplication of basic statistical techniques such as the mean, standard deviation, anad variance to the low income and middle income cohort data from the previous year would provide information for setting realistic criteria for minimum mastery appropriate for the population under study.

IMPLICATIONS FOR PRACTICE

There are several important implications for those involved in measuring school effectiveness that emerge from our research. First, the level of aggregation is a critical issue in assessing school effectiveness. As has been

stated in the literature, data aggregated at the school level often masks the performance of pupils functioning at either end of the scale. The results of our study verify that even a relatively homogenous population is not completely devoid of pupils functioning below their expectancy level. For this reason, algorithms using mean scores with data aggregated at the school level should not be used.

Second, the type of aggregation is an important element to be considered in measuring school effectiveness. Data may be aggregated at the cohort level based on socio-economic status, race, or achievement. Edmonds' concern with the equitable distribution of goods and services seems an appropriate guideline in decisions regarding the level of aggregation and the criterion for selecting cohorts. Practitioners and researchers faced with these choices can determine the appropriate type of aggregation by asking: which component of the population is not getting an opportunity to develop to the fullest potential? Our research suggests that the most equitable measure of school effectiveness is based upon achievement cohorts. Using cohorts based upon achievement allows the school to focus on raising all pupils to grade level standards regardless

of race, social class or ethnicity.

A third issue which must be clearly addressed is the standard against which performance is judged. Our study has revealed that in practice the standards that are used vary greatly. Some algorithms use external norms while others use internal standards (see Figure 1). We found problems with both internal and external standards. Algorithms that rely on an increasing percentage of students achieving at the 30th percentile (internal norm - comparison with past performance) break down as the percentage approaches 100%. Some algorithms rely on comparison with city-wide means. This is an external standard although the school's achievement patterns affect the city-wide mean against which they are measured. A similar study, conducted by Richard Silverman (1984) in a suburban town with several elementary schools, found that a school might be declared effective one year and ineffective the next because the city means fluctuated while the school mean held constant. Both studies also demonstrated that the 30th percentile as a measure of minimum mastery was too low for typical suburban elementary schools. The 50th percentile or higher would be more appropriate.

Finally, our analysis supports the notion that "point-in-time" measures present only a "snap-shot" view of a schools' overall effectiveness. Only three of the six algorithms tested produced consistent results when used to analyze data over three successive years or cohorts of entering students. Of the three that did produce consistent results, two declared the school to be effective and the other declared it ineffective. Consequently, we would caution practitioners to look at achievement data over several years or cohorts and not to make judgments about their schools based on a single application of an algorithm.

CONCLUSIONS

The results of this study have clearly demonstrated that the algorithms for measuring school effectiveness can be used to assess the performance of pupils in suburban elementary schools which are relatively homogeneous in terms of race and socio-economic status. The algorithm developed by Clauset and Gaynor (1982) has proven to be more discriminatory in identifying pupils for whom the school is ineffective. Practitioners and/or researchers should consider using this model, which is based on achievement cohorts whose

performance is compared to grade level standards over a six year period, to measure school effectiveness.

While the focus of this study has been the use of standardized achievement scores as a measure of school effectiveness, there is a temptation that must be warned against. The temptation is to assume that pupil performance can be analyzed in isolation and improvement programs implemented solely on this basis. On the contrary, the research suggests that change to greater academic effectiveness requires a comprehensive approach. Standardized achievement scores represent only one piece of the school effectiveness puzzle.

BIBLIOGRAPHY

Brookover, Wilbur. Why Do Some Urban Schools Succeed? Book Review in Harvard Educational Review. 1981, 51, 139-144

Clauset, Karl, & Gaynor, Alan. "A Systems Perspecive on Effective Schools". Educational Leadership, 1982, 40(3), 54-59

Clauset, Karl, & Gaynor, Alan. Improving Schools for Low Achieving Children: A System Dynamics Policy Study. Paper presented at American Education Research Association Annual Meeting, New York, New York, March, 1982. (ERIC Document Reproduction Service No. ED 214 243)

Dorman, G. Middle Grades Assessment Program: User's Manual. Center for Early Adolescence, Chapel Hill, North Carolina, 1981.

Edmonds, Ronald. "Programs of School Improvement: An Overview". Educational Leadership, 1982, 40(3), 4-11

Gauthier, William J. Connecticut Perspectives on Instructionally Effective Schools. Paper presented at the Annual Meeting of the American Educational Research Association. New York, March, 1982

Kean, Michael H. Issues in Identifying Effective Schools. Paper presented at the National Assessment Conference on Large Scale Assessment, Boulder, Colorado, June, 1982. (ERIC Document Reproduction Service No. Ed 222 571)

McCormack-Larkin, Maureen, & Kritek, William J. "Milwaukee's Project Rise". Educational Leadership, 1982, 40(3), 16-21

Miles, Matthew A., Farrar, Eleanor, & Neufield, Barbara. The Extent of Adoption of Effective Schools Programs Washington, D.C.: Department of Education, 1983. (ERIC Document Reproduction Service No. ED 228 242)

Rowan, Brian, Bossert, Steven T., & Dwyer, David C. "Research on Effective Schools: A Cautionary Note." Educational Researcher, 1983, 12(4), 24-31.

Silverman, Richard H. Measures of School Effectiveness: A Comparison of the Major Algorithms. Unpublished doctoral dissertation, Boston University, 1984.

Villanova, Robert. Measuring and Validating the Characteristics of Instructionally Effective Schools in Connecticut. Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, March 1982. (ERIC Document Reproduction Service No. ED 214 979)

Westbrook, John D. Considering the Research: What Makes an Effective School? Washington, D.C.: National Institute of Education, 1982. (ERIC Document Reproduction Service No. ED 223 008)